Study on Linear Array VLSI Processors with Minimum Latency for Pobotics(ロポット用演算遅れ時間最小形リニアアレーVLSIプロセッサに関する研究)

氏名	ソムチャーイ　キッティチャイクンキット Somchai KittichaiKoonkit
授与学位	博士（工学）
学位授与年月日	平成6年3月25日
学位授与の根拠法規	学位規則第5条第1項
研究科，専攻の名称	東北大学大学院工学研究科 （博士課程）電子工学専攻
学位論文題目	Study on Linear Array VLSI Processors with Minimum Latency for Robotics （ロボット用演算遅れ時間最小形リニアアレーVLSIプロセッサに関する研究）
指導教官	東北大学教授　亀山　充隆
論文審査委員	東北大学教授　亀山　充隆　　東北大学教授　伊藤　貴康 東北大学教授　樋口　龍雄

論文内容要旨

Chapter 1. Introduction

In order to satisfactorily fulfill many potentail missions and applications, robots must be capable of promptly interacting with the outside environment to capture and recognize changes relevant to task objectives, and quickly respond to unanticipated situations. However, the information processing for such robots consists of various computational intensive tasks exhibiting vastly different characterristics, including kinematics, dynamics for manipulator control, trajectory planning, sensor signal processing and human interface. Moreover, sensory feedback is essential to realize the information processing, so that robots can operate to respond unanticipated situations or unpredictable changes of outside environment. By the sensory feedback, next sample inputs in the information processing can be obtained only after robots actually move in the real world. In order to realize the information processing in high speed, it is essential to reduce the latency, i. e, the time interval between the input and the output of data. The most promising answer to reduce the latency is the use of special-purpose VLSI processors for each intensive task in robot information processing, so that we can fully exploit the potential of VLSI technology together with parallel processing techniques, which

is the best suited to the specific computation.

In robot information processing, there are many recursive computations composed of multiple stages of simple operations, and only the communication between the operations in the neighboring stages is sufficient. The multiple matrix multiply-addition for robot control, and the FFT computation usually used in robot image recognition are the examples of this type of computation. Based on the multi-stage data flow of the recursive computation, a linear array VLSI processor where identical processing elements (PEs) are connected linearly is the most natural parallel structure to reduce the latency of the computation. Moreover, the local communication structure of the linear array is also very efficient to exploit the potential of the VLSI technology to reduce the latency of the recursive computation. Thus, the minimization of the latency based on the boundary condition of neighboring communication of the linear VLSI processor becomes very important for robotics.

In this dissertation, we propose linear array VLSI processors with minimum latency for robotics. The latency is minimized based on the newly proposed of *immediate output-generation* in both the linear array scheduling level where PEs are considered as basic blocks, andthe PE architecture level. By the performance evaluation, it is demonstrated that, the latency of the proposed linear array VLSI processors for robot inverse dynamics computation and FFT computation are both less than $1 / 100$ of that of the conventional linear arrays based on general-purpose VLSI processors.

Chapter 2. Basic Considerations on the Design of Linear Array VLSI Processors for Robotics

A linear VLSI processor consists of identical processing elements (PEs) connected linearly. The multi-stage recursive computation to be performed on the linear array is represented by a data-dependence graph, where nodes and arcs are corresponding to the basic operations and data dependencies, respectively. For simplicity, the recursive computation performed by the linear array is restricted to the computation where the communications between all adjacent stages are the same shuffle communication, because the computation is regular and includes many practical recursive algorithms for robotics such as the multiple matrix multiply-addition and the FFT computation.

In the design of linear array VLSI processors, the latency is minimized in both the linear array scheduling level and the PE architecture level. In the linear array scheduling level, the PE is treated as a basic block capable of performing one node in a single clock time, and the latency is minimized without considering the architectural structure inside the PE. Bacause the data flow of the processing in the linear array is corresponding to the relay race, the concept and the algorithm of the scheduling of immediate output-generation for the

minimum-latency computation is proposed. The immediate output-generation is realized using the recursive reshuffle of the input data based on the concept of *immediate input-transfer*. That is,the nodes to be performed at each dtage are reshuffled recursively such that all the input data of the same node of the next dtage are immediately transferred, and then each PE immediately generates its output data without extra delay based on unnecessarily waiting the input data. Using the proposed scheduling, we can simply obtain the minimum-latency schedules of the multiple matrix multiply-addition computation for robot control, the FFT computation for the correlation computation in robot vision and so on.

After we minimize the latency of the linear array scheduling level, the latency is also minimized by the same concept of immediate output-generation in the PE architecture level. In order for the PE to immediately generate its output data to the next stage, we introduce the maximum spatial parallel processing based on the data-dependence graph inside the node. In conventional spatial parallel processing, parallel data transfer with reasonable hardware is usually a difficult and important issue. In order to overcome this problem, we propose the parallel data transfer with efficient hardware based on the use of multiple local memories connected according to the scheduled data-dependence graph obtained in the proposed *immediate output-generation scheduling*.

Chapter 3. Design of a Linear Array VLSI Processor for Robot Inverse Dynamics Computation

This chapter discusses the design of linear array VLSI processors for inverse dynamics computation which is one of the most important computations in robot control. First, we propose the multiple matrix multiply-addition based recursive algorithm for the inverse dynamics computation. The algorithm is very suitable for the linear array VLSI processor because all the computation can be attributed to multiple matrix multiply-additions.

Based on the immediate output-generation scheduling proposed in Chapter 2, it is shown that the total latency required for the inverse dynamics computation can be minimized based on the concept of *odd-even alternative scheduling* where only two types of regular scheduled data-dependence graphs are sufficient for scheduling all PEs in the linear array. Since the node or the basic operation in the multiple matrix multiply-addition is multiply-addition, and since the matrix size necessary in the inverse dynamics computation is only 4 by 4, for the immediate output-generation of the PE, we utilize 4 multiply-adders to construct a single PE. In order to minimize the latency of the linear array with efficient hardware, the 4 multiply-adders in the PE are regularly connected by simple data switches such that the executions according to the two types of the scheduled data-dependence graphs in the *odd-even alternative scheduling* are directly realized. Parallel data access is also efficiently realized

based on the decomposition of the memory into multiple local memories regularly connected based on the scheduled data-dependence graphs.

Fig. 1 shows the layout of the PE. It is demonstrated by the layout evaluation that the proposed PE can be easily implemented in one chip using the present state-of-art VLSI technology. As shown in Table 1, the latency of the proposed linear array is respectively 1/3400 and 1/400 less than the conventional SIMD linear arrays constructed by general-purpose VLSI processors and digital signal processors. In fact, the performance regard to the latency of the proposed linear VLSI processor is the highest in all the architectures reported so far.

Chapter 4. Design of a Minimum-Latency Linear Array FFT Processor for Robotics

This chapter discusses the design of linear array VLSI processors for FFT computation which is often used for high-speed correlation computation for image recognition in robot vision. The FFT computation based on double linear array structure is proposed. The Fourier transforms of the real and imaginary parts of the input data are independently performed in different limear arrays, so that the computation amount in each linear array can be reduced by half based on the introduction of the real-valued FFT computation. Since no communication between the linear arrays is necessary, the latency of the proposed double linear array structure can be reduced by half compared with that of the conventional single linear array VLSI processor.

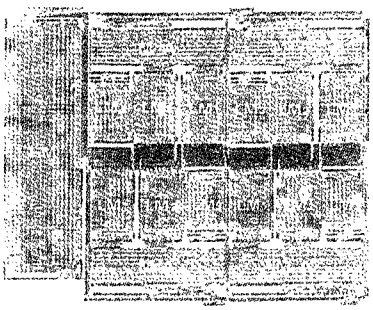
By introducing the mmediate output-generation scheduling proposed in Chapter 2, the minimum-latency schedule for the FFT comutation of each linear array can be easily obtained. In order to realize the immediate output-generation in the PE archiecture level, the PE with spatial parallel structure is designed based on the data-dependence graph of the node (butterfly operation) in the FFT computation. Parallel data transfer with efficient hardware is realized using multiple local memories connected based on the scheduled data-dependence graph. In order to support sufficiently wide dynamic range in the FFT computation, the floating-point PE is desirable. However, in the straightforward design using individual floating-point arithmetic modules connected based on the data-dependence graph, the latency is very large bacause multiple bit alignments and normalizations are necessary in each arithmetic modules, and because multiple stages of carry-lookahead additions (CLAs) are performed for the multiple-input additions in a single butterfly operation. To overcome this problem, we propose the new floating-point module capable of performing the butter-fly operation with less than half latency of the straightforward design. We merge all the arithmetic modules into one floating-point modules so that only one bit alignment and one normalization are

sufficient for all the butterfly operation. Moreover, we propose the CSA-based multiple-input adders to minimize the number of stages of the CLA in the critical delay path.

Fig. 2 shows the layout of the PE. It is demonstrated by the layout evaluation that the proposed PE with 32-bit floating point data format can be easily implemented in a single chip using the present state-art VLSI technology. As shown in Table 2, the latency of the proposed linear array is respectively 1/3500 and 1/600 less than the conventional SIMD linear arrays constructed by generalpurpose VLSI processors and digital signal processors. In fact, the performance regard to the latency of the proposed linear VLSI processor is the highest in all the architectures reported so far.

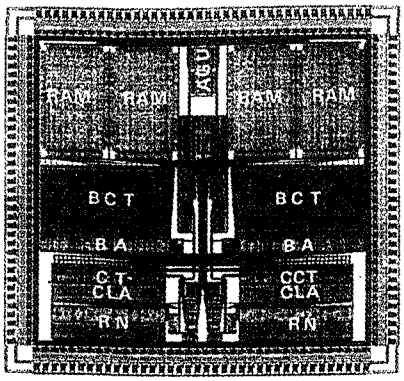
Chapter 5. Conclusions

This cahpter deals with the overall conclusions and some additional comments for the future directions of this study.



MA : Multipy-Adder
LM : Local Memory
IM : Instruction-Memory

Fig. 1. Layout of the PE of the Inverse Dynamics Processor.



AGU : Address Generation Unit, CT : 5-input CSA Tree,
BCT : Booth CSA Tree, CCT : Complement & CSA Tree,
BA : Bit Alignment, RN : Round & Normalize

Fig. 2. Layout of the PE of the FFT Processor.

Table 1.
Comparison of the Latencies of Various Linear Array Processors for the Inverse Dynamic Computation (6 −DOF Arms)

	Proposed Linear Array	Transputers Based SIMD Linear Array	DSPs Based SIMD Linear Aryay
Latency [μs]	3	11525	1382

Table 2.
Comparison of the Latencies of Various Linear Array Processors for the FFT Computation (1024 Point)

	Proposed Linear Array	Transputers Based SIMD Linear Array	DSPs Based SIMD Linear Aryay
Latency [μs]	13	45568	8449

審 査 結 果 の 要 旨

ロボットエレクトロニクスにおいては，センサフィードバックによる実世界からのデータフローが存在するため演算遅れ時間のできるだけ小さい高性能プロセッサチップファミリの系統的設計法が望まれている。

著者は，リニアアレープロセッサ構造が，隣接プロセッサ要素間のみのデータ転送により，規則性のある高集積プロセッサ構成に適合することに着目し，演算遅れ時間が最小となる系統的スケジューリング法を考案した。さらに，それに基づく知能ロボット用高集積プロセッサの設計を行い，その有用性を示した。本論文はその構成をとりまとめたもので，全文5章よりなる。

第1章は緒言である。

第2章では，ロボットエレクトロニクスにおける演算遅れ時間最小形リニアアレー構造プロセッサ開発の意義を述べている。また，瞬時出力応答という概念に基づき，データ依存グラフを利用した演算遅れ時間最小化のための統計的スケジューリング法を考案し，その理論的考察を行っている。

第3章では，知能ロボット制御において重要となる逆動力学演算用VLSIプロセッサの構成法を提案している。繰返しラグランジェと呼ばれる規則的アルゴリズムに対して演算遅れ時間が最小となる最適スケジューリング法を適用した結果，奇数段目と偶数段目でタスク割当てが交互に異なる奇偶交互スケジューリングを見出している。さらに，これに適合するVLSIプロセッサチップを設計し，市販の制御用プロセッサによる並列処理と比較して，少なくとも3桁程度の性能向上が達成されることを明らかにしている。これは重要な成果である。

第4章では，相関演算などに利用されるロボットビジョン用FFTプロセッサの構成法を示している。まず，複素数バタフライ演算を2個の独立した実数バタフライ演算に分割可能であることに着目したダブルリニアアレー構造を考案している。さらに，最適スケジューリングによる得られたデータフローに基づき，浮動小数点演算機能を有するプロセッサ構成法を提案している。高速浮動小数点演算回路など種々のレベルの高速化を駆使することにより，信号処理プロセッサによる並列処理と比較して，少なくとも2桁程度の高性能化が達成されることをレイアウトに基づく性能評価により実証している。これは，実用上有用な成果である。第5章は結言である。

以上要するに本論文は，ロボットエレクトロニクスシステムにおいて有用となる演算遅れ時間最小形リニアアレー構造高集積プロセッサの統計的設計法を確立したものであり，電子工学および情報工学の発展に寄与するところが少なくない。

よって，本論文は博士（工学）の学位論文として合格と認める。